PATENT APPLICATION

5757-00601
Meetrix

"EXPRESS MAIL" MAILING LABEL
NUMBER EL990144362US
DATE OF DEPOSIT MARCH 9, 2004
I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C.F.R. § 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE COMMISSIONER OF PATENTS, MAIL STOP PATENT APPLICATION, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450

Derrick Brown

MEDIA BASED COLLABORATION USING MIXED-MODE PSTN AND INTERNET NETWORKS

invented by:

Thomas A. Dye

Thomas A. Dundon

Assignee: Meetrix Corporation

Attorney Docket No.: 5757-00601

Jeffrey C. Hood
Meyertons, Hood, Kivlin, Kowert & Goetzel, P.C.
P.O. Box 398
Austin, TX 78767-0398
Ph: (512) 853-8800

Priority Claim

This application claims benefit of priority of U.S. provisional application Serial No. 60/453,307 titled "THE METHOD AND PROCESS FOR MEDIA BASED COLLABORATION USING MIXED-MODE PSTN AND INTERNET NETWORKS" filed March 10, 2003, whose inventors are Thomas A. Dye and Tom Dundon which is hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to computer system architectures, and more particularly to audio and video telecommunications for collaboration over hybrid networks.

Description of the Related Art

Since their introduction in the early 1980's, audio/video conferencing systems ("video conferencing systems") have enabled users to communicate between remote sites using telephone lines based on dedicated or switched networks. Recently, technology and products to achieve the same over Internet Protocol have been attempted. Many such systems have emerged on the marketplace. Such systems produce low-frame-rate and low quality communications due to the unpredictable nature of the Internet. Such connections have been known to produce long latencies with limited bandwidth, resulting in jerky video, dropped audio and loss of lip sync.

Therefore, most video conferencing solutions have relied on dedicated switched networks such as T1/T3, ISDN or ATM. Theses systems have the disadvantage of higher cost and complexity and a lack of flexibility due largely to interoperability issues and higher cost client equipment. High costs are typically related to expensive conferencing hardware and dedicated pay-per-minute communications usage. Most often these dedicated communications circuits are switched circuits which use a fixed bandwidth allocation.

In most prior art systems the public switched telephone network (PSTN) is used to transfer audio during conferencing and collaboration with remote parties. It is known that

quality of audio reception is poor over typical prior art Internet protocol (IP) systems. Prior art audio/video conferencing systems which use IP networks for audio and video transport lack the ability to terminate audio to client end systems through both PSTN and IP networks. Thus, it is desirable to achieve a hybrid mix of audio and video data over PSTN and IP based audio/video conferencing to achieve full duplex real-time operation for all conference participants.

Modern voice over IP telephony systems have used the H.323 standard from the international telecommunications union (ITU). The H.323 standard focuses on the transmission of audio and video information through the Internet or switched private networks. Figure 1 illustrates a prior art H.323 system. The block diagram of Figure 1 includes a number of major components, including the general Internet 435, Internet H.323 bridges or gateways 411, telecommunications PSTN 433 (Public Switched Telephone Network), wireless and land-line phone handsets 412/413, standard Internet router 453, an optional gatekeeper 205, a multipoint control unit 203, a standard local area network 457, a voice over IP server running the H.323 protocol 201, and multiple I/O and display terminals 455. Figure 1 is an example of the prior art conferencing system used between hybrid networks connecting the PSTN and Internet. Hybrid networks are used to communicate audio on internal LAN and WAN networks as well as transfer of audio to the existing telephone or PSTN network. While the H.323 recommendation allows for video conferencing, the prior art systems use private switched networks to establish transport that require expensive H.323 bridges between dedicated networks and the PSTN. Each of the components in Figure 3 serves this purpose to achieve audio telecommunications between multiple parties.

Referring again to Figure 1, the components of Figure 1 are interconnected as follows. Prior art technology uses PC or client terminals 455 connected through a local area network 457 to either a data server or a specialized audio/video server 201. The network server 201 contains the application necessary to generate the H.323 network protocol. The data server 201 may be connected to a local gatekeeper 205 that is responsible for management control functions. As known the gatekeeper 205 is responsible for various

duties such as admission control, status determination, and bandwidth management. Data server 201 functions are specified and handled through the ITU-H.225.0RAS recommendations. In addition, management control unit (MCU) 203 is connected to the data server 201. The multipoint control unit of a 203 is required by the eight-step ITU-H.323 recommendation for flexibility to negotiate end points and determine compatible setups for any conference media correspondents. The multipoint control unit 203 enables communication between three or more end points. Similar to a multipoint bridge, the gatekeeper 205 and the multipoint control unit 203 are optional components of the H.323 enabled network. Another useful job of the multipoint control unit 203 is to determine whether to unicast or multicast the audio or video streams. As known by one skilled in the art, these decisions are dependent on the capability of the underlying network and the topology of the multipoint conference. The multipoint control unit 203 determines the capabilities of each client terminal 455 and status each of media stream.

Again referring to Figure 1 a standard network router 453 is connected between the local area network 457 and the Internet 435. At the outer edges of the Internet, "points of presence" are located at multiple end-point or call termination sites. Gateways 411 are used to the transcode the H.323 network information onto the PSTN 433. Standard telephone handsets 413 or wireless phones 412 are connected to the PSTN telephony system.

Figure 2 illustrates the embodiment of the H.323 protocol stack 200, its components and their interfaces to the local area network computers at the network interface 300. The input and control devices 455 along with a local area network 457 of Figure 1 are shown in Figure 2, consisting of the audio input output block 452, the video input and output block 451, the system control unit and data collaboration unit 459. These input devices are largely responsible for the delivery of media data to the H.323 protocol stack 200 shown in Figure 2.

Again referring to Figure 2 the sub blocks of functionality that make up the H.323 protocol stack 200 is described. The H.323 protocol stack consists of an audio codec 211, a video CoDec 213 connected to the audio/video 452 451 input and output blocks. The audio and video CoDecs are responsible for compression and decompression of the audio and video sources. The real-time network protocol component 215 is connected to the audio

video CoDecs and is also responsible for preparation of the media data for transport according to the RTP (real-time protocol) recommendations.

Again referring to the prior art system of Figure 2, the H.323 protocol stack has a system control unit 459 which connects to multiple control blocks within the H.323 protocol stack 200. The system control unit connects to the RTC Protocol block 217 for real-time transport of the control information used to set-up and tear down the conference. The system control unit 459 also connects to the call-signaling units 221 and 219 for call signaling protocols and media stream packetization application used for packet based multimedia communications. The system control unit 459 also connects to the control signaling block 223 used for control of protocols for multimedia communications. Lastly, the H.323 recommendation defines a data collaboration capability as known and outlined in the T.120 data collaboration unit 225.

All of the defined blocks make up the H.323 protocol network interface to the Transport protocol and network interface unit 300 for transport of data through the modem or router 453 to the Internet 435.

Summary of the Invention

The present invention comprises various embodiments which enable audio from standard and wireless telephones systems to be mixed with audio, video and collaboration data resident in IP networks in preparation for transport, preferably over a novel multicasting technique using virtual private networks. In one embodiment, audio data terminating or originating from the PSTN may be multiplexed into open or private IP networks for efficient transport to multiple local or remote client computers. This allows video and audio collaboration clients to talk with remote telephony devices during the process of Multiparty audio/video conferencing.

In an alternate embodiment, without video conferencing, the method may use public networks to transport a multicast enabled IP audio stream during multi-party audio conferencing without the need for a conventional audio bridge device. Audio data is transported in a hybrid network comprising the PSTN and IP network. In this embodiment, a local client initiates a call to the remote telephone or wireless telephony device from a local dial-out application located preferably on the clients' computer. Call set-up is initiated as a series of control packet data transfers to a Voice-over-IP (VoIP) server or PSTN gateway located at some predetermined Internet address on the world-wide-web. Control data packets are transported to the VoIP server via a secure multicast enabled virtual private network. The local client computer compresses the audio data prior to transport to the VoIP server. The VoIP Server uses standard ITU-T, H.323 or SIP audio telephony transport protocol on the primary network connection protocol in preparation for entry to the secondary PSTN. The H.323 or SIP call instantiation is a protocol completed by the VoIP server which requests further transport of the digitized audio stream through a gateway to the public PSTN. In this embodiment, the majority of the audio data in transport over virtual private tunnels is multicast enabled such that the final termination or origination points are geographically close to the local or remote client computers. Once the proprietary data packets are handed off to the VoIP server or remote PSTN gateway, the invention ensures that standard protocols such as H.323 or SIP are used to further process for audio call set-up, tear-down and transport as know by those knowledgeable in the art.

The H.323 or Session Initiation Protocol (SIP) are used for call set-up of the network connections between the Hybrid networks and the remote telephone(s) (PSTN). Once IP network to PSTN call connection is established, compressed digitized audio packet data is grouped into multicast packets and encapsulated for traversal through the open Internet. Transport between the remote PSTN client (Callee) and the Local (Caller) is accomplished with full duplex audio between all audio and video participants within the conference. In one embodiment, compression may be accomplished with a standard audio CoDec such as that specified in the ITU-T G.729 recommendation or with a proprietary audio CoDec as know in the art. Thus, audio data transcoders at the VoIP server may be used to match the expected audio decoders located at the PSTN gateways. The unique process compresses the "Callee" audio data at the local client computer prior to multicast transport to other remote clients and to the VoIP server. This process minimizes the transport bandwidth during the first mile connection to/from the Internet.

In one embodiment, the method for adding a telephone participant to a multi-participant video conference operates as follows. A first message is sent to each of a plurality of multicast appliances over the Internet, wherein the first message comprises a group address which identifies participants. Each of the multicast appliances receives the first message. A plurality of virtual private networks are then established across the Internet between the multicast appliances. As a result, one or more of the participants are able to communicate in the multi-participant video conference. The telephone participant then joins the multi-participant video conference wherein this comprises a first participant contacting the telephone participant; establishing a phone number with a voIP server; the VoIP server communicating with a gateway to call the telephone participant; and the telephone participant participating in the multi-participant video conference.

In one embodiment, the telephone participant participates in the multi-participant video conference as follows: the telephone participant speaking in the video conference; generating digital voice data in response to the telephone participant speaking; transforming the digital voice data into IP packets; transmitting the IP packets containing the digital voice data to the first participant; at a computer system of the first participant, decoding the

IP packets containing the digital voice data to produce the digital voice data; mixing the digital voice data of the telephone participant with digital voice data of the first participant; and providing the mixed digital voice data of the telephone participant and the first participant to the other participants.

The method may further comprise: mixing the digital voice data of the first participant and the digital voice data of the other participants; and providing the mixed digital voice data of the first participant and the other participants to the telephone participant.

In another embodiment, the telephone participant participates in the multi-participant video conference as follows: the telephone participant speaking in the video conference; generating digital voice data in response to the telephone participant speaking; transforming the digital voice data into IP packets; configuring the IP packet with a group address according to a multicast protocol to create a multicast IP packet; encapsulating the multicast IP packet as a unicast packet; transmitting the unicast packet over the virtual private networks across the Internet between one or more appliances; one or more of the appliances determining the multicast data from the unicast packet; and the appliances providing the multicast data to each of the other participants in the group address.

Brief Description of the Drawings

A better understanding of the present invention can be obtained when the following detailed description of the preferred embodiment is considered in conjunction with the following drawings, in which:

Figure 1 illustrates a typical H.323 audio and video conferencing system implemented in accordance with prior art;

Figure 2 illustrates an H.323 protocol stack and its components implemented in accordance with prior art;

Figure 3 illustrates one embodiment of the present invention;

Figure 4 illustrates an embodiment using multicast Protocol;

Figure 5 illustrates the audio and video data flow over hybrid networks; and

Figure 6 illustrates the local client data mixing used in the preferred embodiment.

Detailed Description of the Preferred Embodiment

Incorporation by Reference

The following applications and references are hereby incorporated by reference as though fully and completely set forth herein.

U.S. Application Serial No. 10/446,407 titled "Transmission Of Independently Compressed Video Objects Over Internet Protocol", Dye et al. filed May 28, 2003

U.S. Application Serial No. 10/620,684 titled "Assigning Prioritization During Encode Of Independently Compressed Objects, Dye, et al. filed on July 16, 2003.

International Telecommunications Union Recommendation H.323, Titled "Packet Based Multimedia Communication System." November, 2000

International Telecommunications Union Recommendation H.261, Titled "Video Coding for Audio Visual Services at Px64 kbps."

International Telecommunications Union Recommendation H.263, Titled "Video Coding for Low Bit-Rate Communications" February, 1998

One embodiment of the present invention uses a decentralized model for multipoint conferencing. The multipoint control unit insures communication capability once the media stream is transcoded to the H.323 standard as known. However, this embodiment mixes media streams at each terminal prior to multicast.

Figure 3 illustrates one embodiment of the invention. This embodiment allows audio video and data collaboration information to be securely transferred between a plurality of local and remote clients preferably within a virtual private network. This embodiment provides the ability for a moderator (single member of the conference) to dial out from a desktop computer or terminal (using a novel hybrid network structure) connecting an external telephone user's audio into the audio/video conference. The embodiment integrates full duplex audio, video and data connections between clients conferencing on the Internet and clients conferencing on standard telephone systems. The Internet/PSTN hybrid network is the medium used for transport. Figure 3 depicts the

necessary equipment and protocols to complete the dial out to PSTN network method and process.

Now referring to Figure 3, the voice over IP moderator 401 (call initiator or caller) typically has a number of peripherals used for real input output devices at the desktop. These include a client computing devices such as a PC or other computer 459, a client terminal 455 including a keyboard and mouse for input output control, a standard desktop telephone 457, a video input device or camera 401 and the audio input device, microphone 452. In one embodiment each conference call connected to the Internet will have similar peripheral hardware devices. Figure 3 illustrates a multi-party virtual conference connected over the Internet. Internet clients include audio video client 415, audio video client 418, and audio video client number and 417. In addition Figure 3 shows two possible telephony clients using standard wired 412 or wireless telephone 413 systems. PSTN client #1 412 is connected to a wireless cell-phone that in turn is connected to the global dial network 450 as specified by the PSTN 433. Remote telephony user client 2 413 is connected to a standard telephone handsets 413 which again is connected to the global dial network 450 based on the PSTN 433.

Again referring to Figure 3 the Internet based clients 401 415 418 and 417 are connected through routers or modems 453 preferably in a virtual private network configuration 461. A virtual private network bridge 407 is used to connect local and remote clients together within a secure private network. A local connection from the VPN bridge 407 to the voice over IP server 409 is used to transfer conference audio from any participant on the IP network to any participant in the PSTN. Thus, the voice over IP server 409 is responsible for transcoding audio information from the virtual private network 461 to and from the PSTN gateway 411, thus bridging the PSTN and VPN together.

Figure 4 illustrates one embodiment of the present invention. The system of Figure 4 performs audio transport between multiple client groups who all share the same multicast group address such that audio/video and data may be shared interactively without the need of central servers. Multicast protocol and encapsulated media packets are implemented so that media data may be routed through public or private IP networks without the need for

special hardware and software during the majority of the network transport. Figure 4 shows a system of virtual networks that interconnect as a virtual private network 423. Each VPN tunnel can be connected in a series or star topology between one or more multicasting appliances 449 – 451. One or more central servers or VPN bridge(s) 407 are at the center of the network topology. Multicasting enabled appliances 447, 449, 451, 453, 455 are used at the origination or termination points for audio, video or data (media data) to from the backbone of the transport path. PSTN gateways are used to provide "points of presence" throughout and are responsible for origination or termination of audio data on and off of the PSTN from the IP network topology. Multicast enabled routing allows remote clients to be PC's or PSTN gateways which become "Listeners" of media data. Thus, media data is presented or broadcast onto a network with one or more group addresses. This method uses less bandwidth and reduces latency during transport.

Again referring to Figure 4 PSTN group #1 412 has three analog telephones which are switched into a PSTN gateway and VoIP server 471 which is networked over public or private network connection to a multicast enabled VPN appliance 447. Appliance 447 is connected to a VPN bridge server 407 also by means of a virtual private network. The VPN Bridge 407 is used to authenticate clients, assign multicast IP group addresses to various PC clients and VoIP gateway servers. In addition the VPN Bridge Server 407 may have additional meeting room or conferencing features necessary to carryout a multi-party conference. Connected to the VPN Bridge 407 are various virtual private networks which form network tunnels to one or more other multicasting appliances 449, 451, 453, 455, 457 which connect to one or more PSTN gateways typically located in geographically dispersed areas.

For the purpose of the illustration of Figure 4, PSTN group #1 412 is audio conferencing with PSTN client #3 414 and PSTN client #5 416, each of which are audio conferencing with Audio/Video client group #4 415. In the illustration of Figure 4 each member of audio/video client group #4 share audio with all the clients and video with each other. One example may be illustrated again referring to Figure 4. If telephone client #5 416 is talking, the analog audio is converted from switched network (PSTN) to IP in the VoIP/PSTN gateway 475. The digital IP is routed via Internet to an appliance 455 at the

edge of the network typically co-located with the VoIP/PSTN gateway 475. The appliance has been configured to have a virtual private network creating a tunnel through Internet to appliance 453 which also has Internet based virtual private tunnels to appliance 457 and appliance 447. Audio from PSTN client #5 416 is broadcast from appliance 457 whereby all the audio/video client PC's of group #4 are "listeners" and receive the audio from PSTN client 416 at the same time. Additionally, PSTN client #5's 416 audio is routed over another virtual private network to one or more appliances in this case appliances 447 and 449. PSTN Client group #1 412 are also "listeners" of the multicast group as well as PSTN Client #3 414. Thus, audio is broadcast to multiple audio devices in both IP networks and the PSTN using a unique group address and a virtual private network structure. Interactivity is gained by using the same process no matter who in the group is the broadcaster of audio or video.

Figure 5 shows a more detailed block diagram of the embodiment of the present invention. The moderator client #1 401 initiates the call using the application code running on the voice over IP server 409. Call initiation and call transfer may be accomplished through a VPN tunnel 421 connected to the moderator client 401. Two connections to the Moderator client 1 401 through the VPN tunnel 421 are established. The first connection connects the VoIP conference data for call initiation, set-up and control 405. The second connection 403 through the VPN tunnel connects the conference audio and video 403 between the moderator client 401 and multiple remote clients 415 417 413 connected to the Internet. The VPN tunnel 421 is connected into the VPN bridge 407 which may be located within the Internet 435 at either local or remote sites. As indicated in Figure 5, the VPN bridge 407 is responsible for connecting and establishing the virtual private network used for secure conferencing. In the embodiment of the present invention the VPN bridge 407 bridges all the tunnels for data transfer. Thus, VPN tunnel 421, VPN tunnel 423 and VPN tunnel 425 are on the same virtual private network. Alternate embodiments may include a plethora of tunnels connected to through a single VPN bridge or multiple VPN bridges based on scalability of the system. An additional tunnel containing the conference voice over IP audio and call set-up data 405 is connected to a separate voice over IP server 409. The server 409 is responsible for transcoding the voice over IP audio and call set-up

control 405 in preparation for data transfer across the H.323 network 437. The H.323 network 437 traverses across the Internet to one of many PSTN gateways 411. PSTN gateways 411 form the bridge between the Internet and the public switched telephone network 433. These VoIP gateways are typically located at the local exchange carrier (LEC) in a plethora of individual points of presence throughout the world. Audio telephony calls are terminated at the voice over IP client 413. These termination points may be located throughout the world. Thus, the embodiment shown in Figure 5 allows for the dial-out to standard phones from a client terminal with audio and video capability over IP networks allowing conferencing between multiple remote sites including secure voice over IP audio components over the PSTN.

Figure 6 of the preferred embodiment shows the multiple network domains, the software applications and operating system boundaries and the operations necessary for audio manipulation and transport. It is noted that video accompanies the audio to all conference participants with the exception of the PSTN client 412. For simplicity of illustration, Figure 6 does not show the video conferencing path. The embodiment of Figure 6 includes a local moderator client 401 who is responsible for initiating a dial out for audio conferencing to the PSTN client 412. The local moderator client 401 may also be the initiator of the meeting. In this exemplary embodiment, it may be assumed that the local moderator client 401 has set up the audio video conference with remote audio video clients 418 previous to the dial out for audio conferencing to the PSTN client 412. The local moderator 401 and the remote audio video clients 418 may share audio and video data in a full duplex mode among to all participants with the exception of the PSTN client 412. The PSTN client 412 may share audio from a standard telephone or wireless telephone with all participants in the conference including the local client 401 and remote audio video clients 418. Likewise, the remote audio video clients 418 and the local moderator client 401 may share audio with the remote PSTN client 412. Thus, as indicated in Figure 6 a voice over IP call placed the standard telephone system may bring a remote telephone user into an audio/video conference with multiple remote participants.

A detailed description of Figure 6 follows. It may be assumed in this embodiment that the functions and features of Figure 6 are running on general-purpose hardware using

various software to accomplish the tasks at hand. In alternate embodiments various pieces of Figure 6 may be encompassed in specialized hardware for improved speed performance. Again referring to Figure 6 and starting with the local moderator client 401, the process of call set-up is first performed. The local moderator client 401 uses a computer terminal connected to a local area network that in turn is connected to a wide area network and preferably then connected to a virtual private network 461. The local moderator client 401 is equipped with proprietary software as depicted in Figure 6 to operate as a dial-out to PSTN application. The application interface allows a point and click interface establishing the dial out phone numbers to various possible clients on the PSTN 433. In alternate embodiments "Dial-In" may be used in addition using the same techniques outlined but in a reverse path scenario.

Once the local moderator client 401 has selected the remote PSTN client 412 phone number a point and click on the name initiates the dial-out process where audio information is to be transport across hybrid networks. General tones as known in the art according to the ITT standard are sent from the local moderators computer or terminal to the voice over IP server 409 located somewhere within a global Internet system 435. The voice over IP server 409 may be connected to a virtual private network 461. The voice over IP server 409 may use standard H.323 or SIP network protocol to establish communications as known directly to the PSTN gateway 433. Once the call set-up is complete both the PSTN client 412 and the local moderator client 401 have established a connection. In one embodiment the connection is not established for all the audio participants within the conference at this time. In the embodiment of Figure 6 it is assumed that all the remote audio video clients 418 had previously been in a conference with the local moderator client 401. In alternate embodiments the order at which callers are established may be different. With the foregoing assumption of a conference being established prior to the call-out to PSTN, further definition of the VoIP audio path is specified. The following discloses and further defines the audio paths through three layers of application software 562, 564, 566, including the audio paths through four hybrid network boundaries 510 520 435 and 515.

Starting with the remote client/moderator boundary 510 preceding to the local client voice over IP boundary 520, the Internet interface boundaries 435 and the PSTN telephone

network boundary 515, each of these distinct boundaries makes up the method used to transport audio media in a hybrid mixed network system. Remote client/moderator boundary 510 may be established as a virtual private network for transport of audio and video data between the local moderator client 401 and remote audio/video clients 418. In alternate embodiments the virtual private network may be replaced with either switched dedicated network or standard non-secure IP networks. The local clients VoIP boundary 520 may also be a virtual private network connecting audio from the local moderator client 401 to a local or remote voice over IP server 409. In alternate embodiments the local client voice over IP boundary may be established through switched networks or the open Internet. For security purposes all connections that traverse across the open Internet 435 are preferably secured by the use of encryption running within a virtual private network. Alternate embodiments may exclude encryption and virtual private networks including public non-encrypted information, public Internet interfaces or over private switched networks. Continuing with the description of the Internet interface 435 it is assumed with all the information above the PSTN boundary 515 (as indicated in Figure 6) is information which travels within local client local area networks remote client local area networks or on wide area networks through the Internet. The final boundary for network transport is the PSTN boundary 515. This is the transport interface between the wide area network (Internet) and gateways that transmit data to and from the PSTN system 433.

Again referring to Figure 6 and assuming the PSTN dial out call has been established as known in the art, (preferred to ITU H.323) the following detailed information regarding the audio processing follows. In one embodiment the interface between the conference application boundary 562 and the operating system interface boundary 564 and the voice over to IP application boundary 566 is taken under consideration. Preferably, the operations preformed on the audio occur in real-time to achieve full duplex operation. In alternate embodiments a plethora of alternative methods, operating systems application software, and input and output devices may be used to achieve the same goal as described previously. In one embodiment the operating system sound interface and API boundaries 564 are used for standard audio mixing. The audio from the local moderator client 401 is preferably mixed to be transported both to the PSTN client 412 and remote audio video

clients 418. The conference application boundary 562 is responsible for the application which controls mixing of audio to the operating system sound interface 564. In one embodiment, the operating system sound interface also performs the interface and mixing for the voice over IP application boundary 566. These layers make up the application interface for achieving the operation as described herein. Input from the local moderator client 401 is input to two mixers. First, the moderator audio input 550 is connected to the voice over IP record mixer 568. Secondly, the microphone from the moderator client 401 is also connected to another standard mixture 534. The voice over IP record mixer 568 mixes the audio from the audio decompressors 525 and the local moderator audio 401 in preparation for transport to the voice over IP encoder 522. In addition the local moderator client 401 sends audio to the audio mixture 534 mixing the audio from the voice over IP decoder 524 for output to the conference applications 562 local audio encoder 520. The audio encoder 520 combines the PSTN client 412 audio with the local moderator clients 401 audio then encodes the result for compression of the data in preparation for transport across the VPN network 461. The application software audio encoder 520 delivers both the PSTN client's audio and the local moderator client's audio to remote audio video clients 418.

The local moderator client 401 receives audio from the PSTN client 412, and thus the voice over IP player mixer 569 mixes audio previously decoded by the voice over IP decoder 524 with the audio from the remote client's 418 for presentation to the local speaker 454. All the remote audio video clients 418 hear the audio from the PSTN client 412. The PSTN client 412 transports audio through the PSTN 433 to Internet based voice over IP server 409. The voice over IP server transcodes the audio data into a format suitable for transport onto the VoIP application boundary 566. Figure 6 also depicts how audio data from the remote audio video clients 418 is prepared for transport across a VPN network 461. This audio data is input to the application's local decoders for audio decompression 525 prior to the mixing process. The remote audio video clients 418 audio is mixed with the local moderator client audio 401 in preparation for compression by the VoIP encoder 522. This audio data is then placed in the virtual private network tunnel for

transport to the voice over IP server 409 and onto the gateway for audio presentation to the PSTN, terminating at the PSTN client 412.

Figure 6 outlines multiple application software boundaries used to mix audio between local and remote clients in hybrid data networks as indicated by the multiple protocol boundaries 562, 564, 566. Thus, the embodiment allows enhancements to the ability for audio video conferencing with multiple clients and the added value of dialing out to a remote telephone user located somewhere within the global dial-up network 450. Prior art techniques such as that known in the ITU H.323 recommendations have the compressor 522 and decompressors 524 located within the VoIP server running the H.323 network system as indicated in Figure 2 (audio codec 211). This poses a problem for low bit-rate networks especially when video and audio are already part of the transport data. The present embodiment uses highly compressed audio that is compressed and decompressed at the client computer. Thus, the voice over IP server can be located anywhere within the Internet 435 without concern about the limited bandwidth of the first and last mile. In addition, only a single server is required for multiple conferences. The prior art systems as shown in Figure 1 place at least one or more voice over IP server behind the fire-wall and corporate router for transcoding information to the H.323 network. This requires additional cost when a separate server is needed in each location to run the H.323 standard. The present embodiment does not require a separate server at each site, but instead requires that the desktop computer or terminal compress the data prior to transport.